

SYSTEM AND METHOD FOR SCALABLE MULTI-LEVEL REMOTE DIAGNOSIS AND PREDICTIVE MAINTENANCE

FIELD OF THE INVENTION

[0001] The present invention relates generally to predictive maintenance, and more particularly relates to diagnosing processes and machines at remote locations.

BACKGROUND OF THE INVENTION

[0002] Manufacturing down-time due to machine failure costs industries billions of dollars each year. Several techniques for managing these costs have been developed and are now widely used. These techniques include preventive maintenance and predictive maintenance.

[0003] Time-based preventive maintenance is one of the popular techniques currently employed by the manufacturing industry for reducing the number of unscheduled shut downs of a manufacturing line. In time-based preventive maintenance, components are inspected and/or replaced at periodic intervals. For example, a bearing rated for so many hours of operation is always replaced after a set number of operational hours regardless of its condition.

[0004] Chart 1 shows typical failure probability charts for a variety of components. Curve 1000 illustrates the failure probability of components subject to dominant age-related failure and "infant mortality" (i.e., high initial failure rates decreasing over time to a stable level). Curve 1002 illustrates the failure probability of components having a dominant age-related failure mode only. Curves 1004, 1006 illustrate the failure probability of components subject to failure fatigue. Curve 1008 illustrates the failure probability of complex electromechanical components without a dominant failure mode and electromechanical components that are not subject to an excessive force. Curve 1010 illustrates the failure probability of electronic components (e.g., controllers, sensors, actuators, drives, regulators, displays, Places, computers).

[0005] Time based preventive maintenance decreases failures for components that exhibit a failure probability illustrated in curves 1000, 1002. These components, which comprise a low percentage of approximately four to six percent of installed equipment, include complex mechanical equipment subject to premature failures (e.g., gearboxes and transmissions) and mechanical equipment with a dominant age-related failure mode (e.g., pumps, valves, pipes). Preventive maintenance does not decrease or increase failures for components that exhibit a failure probability similar to the failure probability illustrated in curves 1004 - 1008. However, if some other component is disrupted during the maintenance, the failure rate of these components actually increases with time based preventive maintenance.

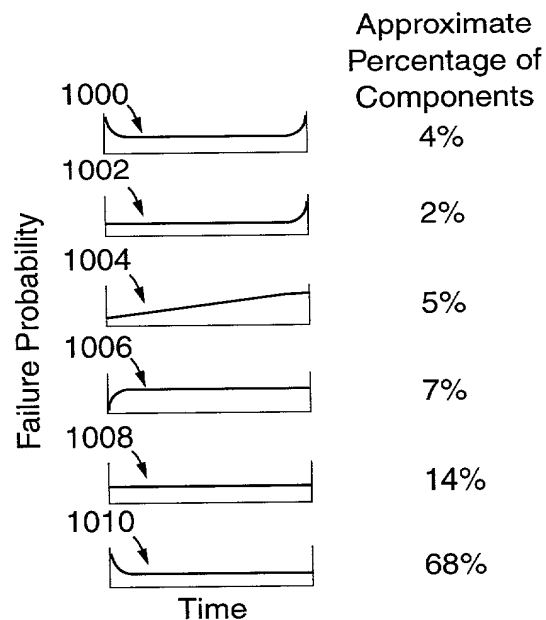


Chart 1

[0006] Time based preventive maintenance actually increases the failure rate of electronic components by prematurely shutting down a manufacturing line for scheduled maintenance and introducing "infant mortality" in what is an otherwise stable system. Curve 1100 in chart 2 illustrates the increased failure probability due to "infant mortality" when electronic components are replaced due to preventive maintenance and curve 1102 illustrates the failure probability with no preventive maintenance performed.

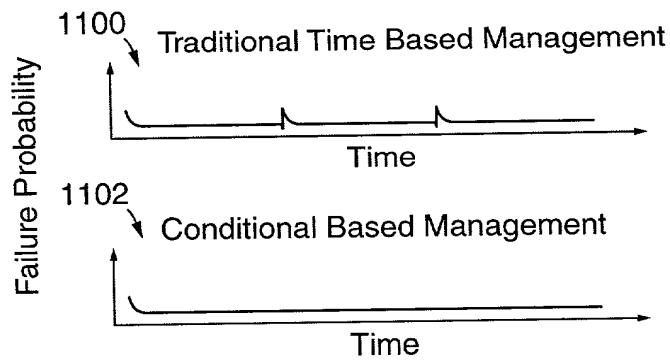


Chart 2

[0007] The manufacturing industry has recognized these and other problems with preventive maintenance, but the alternatives are expensive. One of these alternative techniques is predictive maintenance. In its most simple form, predictive maintenance monitors the condition of operating parameters on a machine over a period of time. Predictions are generated of when a component should be replaced based on detected changes in the operating parameters. The changes can also be used to indicate specific faults in the system being monitored. Techniques for predictive maintenance that are available today, however, are either poorly matched to the particular circumstances and, therefore, less than completely effective or they are so expensive as to be prohibitive in all but the most expensive manufacturing settings.

[0008] Predictive maintenance systems have had only a limited acceptance by the manufacturing industry. It has been estimated that these systems are being used today in less than one percent of the total maintenance market. Many predictive maintenance systems are expensive, require local experts, and are often unstable or unreliable. These systems require continuous monitoring of operating parameters and conditions. This continuous monitoring results in an enormous amount of data that, in turn, requires significant processing power. As a result, predictive maintenance is often cost-prohibitive. Due to the expense of the installation and maintenance of these predictive systems, manufacturers either limit the number of systems installed in a manufacturing site, limit the number of components at the site that are monitored, or perform time sampling of components instead of continuous monitoring. The reduced monitoring reduces the effectiveness of the system and ultimately results in its unreliable performance.

[0009] Other problems that are a direct or indirect result of manufacturer's efforts to reduce cost have been encountered. One problem is when on-site technicians periodically collect machine condition data. The periodic manual collection of data is expensive and results in discontinuous monitoring. The discontinuous monitoring leads to an increased failure rate of machines due to machines failing before a failure is diagnosed due to the lag in time. Additionally, the sensors used to collect the data may not be permanently mounted, which results in the sensors being located at a slightly different location each time data is collected. As a result, any difference between data measurements may be due to the change in location of the sensors and not the machine being monitored.

[0010] Another problem is with the use of experts that analyze the data. Locally based experts may be difficult to find. Transmitting all data to an expert for analysis requires bandwidth. Additionally, experts are expensive and often become a bottleneck in the process.

[0011] Another problem encountered is when sophisticated local modeling and signal analysis tools are used. The configuration of these tools requires a skill level that is not always available. Additionally, the model becomes obsolete when a minor change to the machine is made, requiring re-generation of a new model. Conversely, centralized signal analysis can become overloaded as additional data is received for analysis.

[0012] Another problem is that present systems lack scalability. These systems are typically designed for a specific implementation and become overloaded as the number of systems being monitored increases. These systems also require complex customization for each new system.

[0013] Another problem is the reliability of sensor signals used to monitor the system. It has been estimated that fifty percent of monitoring problems are a direct result of sensor failure, sensor obstruction (e.g., oil, dust, or other particles), and severed or damaged sensor cables. The present monitoring systems typically do not monitor the health of the sensors used to monitor the system.

BRIEF SUMMARY OF THE INVENTION

[0014] The invention provides a method for remotely monitoring and diagnosing operations of a device, machine, or system (hereinafter called "machine") and for performing predictive maintenance on a machine. A signal model of the machine is created based on sensed signals during normal operation of the machine. Signals representative of the machine's operating and condition parameters are sensed and compared to the signal model locally maintained proximate to the device in order to detect anomalies. Once an anomaly is detected, information describing each anomaly is transmitted to a location remote from the machine. The information is diagnosed at the remote location. The signal model is adapted to work with the remaining sensors if a failed sensor is detected.

[0015] The diagnosis includes an initial analysis of the information by diagnostic tools maintained at the remote location. The diagnostic tools include a library of patterns comprising information describing systemic anomalies and a library of patterns comprising information describing component anomalies. The information is compared to patterns in the library describing systemic anomalies and component anomalies for a match. If a match is found, a diagnosis is made.

[0016] If the initial analysis fails to provide a diagnosis, a subsequent analysis of the information by diagnostic tools maintained elsewhere is performed. A final analysis by a team of humans aided by a collaborative environment is performed if the initial and subsequent analyses fail to provide a diagnosis. The diagnosis of the anomaly is reported to a location capable of attending to repair of the machine. Each new diagnosis is added to the appropriate pattern library for analysis of future anomalies, which improves the diagnostic capability of the system.

[0017] Other features and advantages of the invention will become more apparent from the following detailed description when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The accompanying drawings incorporated in and forming a part of the specification illustrate several aspects of the present invention, and together with the description serve to explain the principles of the invention. In the drawings:

[0019] FIG. 1 is a block diagram generally illustrating an exemplary environment in which the present invention operates;

[0020] FIG. 2 is a flow chart of a method of diagnosing failures of components in accordance with the present invention;

[0021] FIG. 3 is a block diagram of an exemplary end user plant in which part of the present invention operates according to one embodiment of the present invention;

[0022] FIG. 4 is a block diagram of an embodiment of a local detector in accordance with the present invention;

[0023] FIG. 5a is a flow chart of an exemplary process performed in level 202 of the flow chart of FIG. 2;

[0024] FIG. 5b is a flow chart of an exemplary process performed in level 204 of the flow chart of FIG. 2;

[0025] FIG. 5c is a flow chart of an exemplary process performed in level 206 of the flow chart of FIG. 2;

[0026] FIG. 5d is a flow chart of an exemplary process performed in level 208 of the flow chart of FIG. 2; and

[0027] FIG. 6 is a block diagram illustrating the step of auto-configuring a communications link in accordance with the present invention;

[0028] While the invention will be described in connection with certain embodiments, there is no intent to limit it to those embodiments. On the contrary, the intent is to cover all alternatives, modifications and equivalents as included within the spirit and scope of the invention as defined by the appended claims.

DETAILED DESCRIPTION OF THE INVENTION

[0029] Turning to the drawings, wherein like reference numerals refer to like elements, the invention is illustrated as being implemented in a suitable operating environment. The invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[0030] Figure 1 illustrates an example of a suitable operating environment 100 in which the invention may be implemented. The operating environment 100 is only one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of the invention. The operating environment 100 includes one or more end users 102 in communication with an OEM server 108 via a network 106. Each end user 102 comprises a location where one or more machines or devices are located. For example, an end user 102 may be a manufacturing plant, a remote station or machine, a business, a home, a vehicle, or any other place where reliability of equipment is a concern. The end users 102 are connected to the network 106 via proxy/gateways 104. The network 106 in one embodiment is the Internet. Alternatively, the network 106 may be a virtual private network, a dedicated network, a public switched network, a wireless network, a satellite link, or any other type of communication link.

[0031] The OEM servers 108 communicate with each other in a peer-to-peer network 110. Those skilled in the art will recognize that the network 110 may be other type of networks such as a virtual private network, a dedicated network, or any other type of communication link. A directory server 112 maintains a list of all OEM servers 108 and, as described hereinbelow, is used to aid OEM servers find other OEM servers. The directory server 112 also communicates with the expert network server 114. The expert network 114 maintains a list of available experts located in a collaborative network 116 that can be used to solve particular problems.

[0032] Turning now to FIG. 2, the operating environment 100 of the present invention has four levels. Level 202 includes the end user 102 and proxy/gateway 104. The equipment 300 being monitored is located in the end user location (see FIG. 3). A detector 302 that

monitors the machine 300 with sensors 304 is located proximate to the machine 300. Each detector 302 is in communication with the proxy/gateway 104 via a wireless LAN 306 and sends data to an OEM server 108 if a problem is detected. Alternatively, the detector 302 communicates with the proxy/gateway 104 through a powerline carrier for signal transmission.

[0033] Returning to FIG. 2, level 204 includes an OEM server 108. The OEM server 108 hosts an expert system that analyzes the data received from the detector 302 and diagnoses the problem. If the OEM server 108 is unable to diagnose the problem, the data is sent to other OEM servers 108 that are selected by the directory server 112 in level 206.

[0034] Level 206 includes the OEM servers 108 in the network 110 and the directory server 112. When data is received from the OEM server 108 in level 2, the selected OEM servers 108 attempt to diagnose the problem. The diagnosis and solution are returned to the OEM server 108 in level 204. If the selected OEM servers 108 are unable to diagnose the problem, the data is sent to the expert network server 114.

[0035] Level 208 includes the expert network server 114 and the collaborative network 116. In level 208, experts are chosen to diagnose the problem. The experts are located throughout the world and the collaborative network 116 allows the experts to diagnose the problem without having to travel from their home locations. When the problem is diagnosed and solved, the solution is returned to the OEM server 108 in level 204.

[0036] Now that the overall system has been described, further details of the detector 302 and the process used to diagnose and solve a problem will be described. Although not required, the invention will be described in the general context of computer-executable instructions, such as program modules, being executed by a computer. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Moreover, those skilled in the art will appreciate that the invention may be practiced with other computer system configurations, including hand-held devices, multi-processor systems, embedded devices, microprocessor based or programmable consumer electronics and consumer appliances,

network PCs, minicomputers, mainframe computers, and the like. For purposes of illustration, the invention will be described in terms of monitoring a machine. Those skilled in the art will recognize that the present invention can be used on any type of installation or device where reliability is a concern and in any location (e.g., inside an installation, in an automobile or truck, in an outdoor environment, etc.)

[0037] Turning now to FIG. 4, a block diagram of an embodiment of the detector 302 is shown. The detector 302 includes a power supply module 400, analog sensor input module 402, reset/relearn button 404, indicator 406, communication module 408, and a central processing unit (CPU) 410. The primary functions of the detector 302 are sensor data collection and buffering, data transformation using fast Fourier transforms (FFTs) or other transformation techniques, statistical model generation, real time model data calculation, real time decision making, sensor health monitoring, communication with the proxy/gateway 104, and local indication of machine status.

[0038] The power supply module 400 provides power to the other components of the detector 302. The analog sensor input module 402 receives and processes signals from sensors 304 mounted on or proximate to the machine 300 being monitored. The sensors 304 are connected to the analog sensor input module 402 by point-to-point wire connections, a sensor bus, or a wireless connection. The sensors 304 are used to monitor the machine's operating and condition parameters. The operating and condition parameters include parameters such as vibration, speed, rotor position, oil temperatures, inlet and outlet temperatures, bearing temperature, pressure, power draw, flow rates, harmonic content, etc. The sensors 304 include vibration sensors, temperature sensors, speed/position sensors, electrical parameter sensors (e.g., voltage and current), pressure sensors, flow rate sensors, and status inputs. The analog sensor input module 402 performs filtering and other signal conditioning when necessary. For example, vibration sensor signals typical require high pass filtering to filter out undesirable low frequency noise and at least one gain stage to optimize signal levels. Those skilled in the art will recognize that many functions of the analog sensor input module 402 may be integrated into individual sensors as sensor technology improves.

cost. Once the hierarchy is established, the non-permanent connections such as the dial-up modem are disconnected to reduce cost (step 610).

[0042] Returning now to FIG. 5a, the detector 302 generates a statistical signal model for the machine 300 (step 504). This step is performed by the detector 302 entering into a learning mode to learn how the sensor signals correlate with each other during normal operation. The detector 302 enters into the learning mode during installation and start-up and whenever the detector 302 is commanded to enter the learning mode. The command to enter the learning mode is transmitted remotely or locally. The reset/relearn button 404 is pressed to enter the learning mode locally. The remote command is received through the communication module 408. In the learning mode, the detector 302 obtains representative data (i.e. training data points) from the sensors 304 for a predetermined user-configurable number of sampling periods (e.g., sample ten sensors at a 5 kHz rate for sixty seconds). The detector 302 then fits the best reference curve(s) through the training data points as known in the art to generate the statistical model. Those skilled in the art will recognize that there are a wide variety of methods that can be used to fit the curve and a wide variety of optimization points that may be chosen. Additionally, there are a number of different types of curves that may be used (e.g., higher order curves such as second order, third order, fourth order, etc. or multiple-segment linear curves). As statistical modeling techniques improve or are developed, the detector 302 is updated with the new/improved techniques.

[0043] After the model has been generated, the detector 302 monitors the operation of the machine 300. In this phase of operation, the detector 302 obtains the processed data and performs an FFT or other transformation algorithm on the data (step 506). The detector 302 has enough memory to hold a working data buffer for the processed data (i.e., the sensor data in which filtering, amplification, integration, A/D conversion and similar operations have been applied). For example, in one embodiment, five minutes of data for ten sensors with 16 bit resolution at a 5 kHz sampling rate requires a storage capacity of approximately 30 MB. The detector 302 also maintains an incident archive and a context archive. Each archive contains 120 FFT images of all sensor data for relevant high sampling rate sensors. For example, accelerometers or current sensors would be part of the FFT images but temperature sensors would not because a single value for temperature would be sufficient. The incident

archive contains one FFT per minute for two hours. The incident archive is cyclically rewritten so that after two hours, each data entry is deleted. Before deletion, one FFT per hour (i.e., two FFTs from the entire incident archive) is moved into the context archive and kept for five days (i.e., 120 hours). The data in the incident archive and context archive is not analyzed by the detector 302. In the event that sensor data does not fit the model as described below (i.e., an anomaly), the incident and context archives are transmitted to the OEM server 108 in level 204, where it is compared to the systemic pattern library. In the event that human experts are needed to solve a problem the data in the incident and context archive is transmitted to level 208 and utilized by human experts. The memory required for each archive is approximately 240 kB. It should be noted that the size (i.e., number of samples) and sampling rate of the incident and context archives can be reconfigured.

[0044] The detector 302 compares the actual sensor data to the statistical model to determine if the sensor data changes relative to the statistical model in a similar manner (step 508). This step is performed by calculating the distance between the model reference curve and each actual data point. These distance points are analyzed over a period of time. If the distance remains small and random (i.e., the sensor data fits the model), the machine 300 is operating normally (step 510) and steps 506 and 508 are repeated. A signal is sent periodically to the OEM server 108 to indicate that the machine operation is normal. If the distance does not remain small and random (i.e., the sensor data does not fit the model), the detector 302 transmits the sensor data to the OEM server 108 (step 512), provides a visual or audio alert by changing the status of the indicator LED 406, and continues monitoring the machine 300 by repeating steps 506 -512. The sensor data is compressed prior to transmission (for faster and more cost-effective transmission) and sent to the OEM server 108 via the proxy/gateway 104. If the anomaly persists, the detector 302 periodically transmits transformed data in batches to the OEM server 108 in order to avoid OEM server saturation and excessive transmission costs.

[0045] In an alternate embodiment, the detector 302 does not fit a reference curve through the training data points. The detector 302 selects a relevant subset of the training data that is representative of normal machine operation and compares the actual sensor data to the subset of training data as described above. The distance between the selected training

data points and actual data points is used and analyzed over a period of time. In a further alternate embodiment, virtual sensors are created for a select number of real sensors by maintaining a weighted moving average of sensor data and comparing the actual sensor data to the weighted moving average over a period of time. Those skilled in the art will realize that other alternatives may be used. The alternatives must meet the criteria of balancing robustness, accuracy, and fast model generation using standard processors.

[0046] During operation, the detector 302 also monitors the health of sensors 304. The health is monitored by first calculating an estimated sensor signal from other sensor signals and the statistical model. The difference between the estimated sensor signal and actual sensor signal is compared. If the difference is not small and random, an alert is provided that the sensor has failed. The failed sensor is excluded from further model calculation until it is repaired or replaced. After a failed sensor has been repaired or replaced, the detector 302 waits until it enters the learning mode before it uses the sensor in the model calculation. The sensor health monitoring is repeated periodically for each sensor at an appropriate period of time. For most sensors, a time period of once per second is adequate.

[0047] Turning now to FIG. 5b, the OEM server 108 in level 204 receives the sensor data transmitted by the proxy/gateway 104 and decompresses the data. The OEM server 108 hosts an expert system that has a component pattern library and a systemic pattern library. The OEM server 108 or its components (e.g., expert system) may be integrated with other components of an installation, including operating as a software object on any processor or distributed processors having sufficient processing capability. The component pattern library contains known component specific failure patterns. For example, the component pattern library may contain failure patterns for ball bearings, motors, gearboxes, cams, etc. The systemic pattern library contains systemic patterns as diagnosed by human experts. This library is updated each time an expert identifies and classifies a new pattern. The patterns can be characterized either as normal operation or as a specific failure situation. The expert system automatically generates a model of a machine's systemic behavior each time a pattern is added to the systemic pattern library.

[0048] The OEM server 108 compares the sensor data with known systemic patterns in the systemic pattern library using a model of systemic behavior (step 520). If there is a match between the sensor data and a specific failure pattern in the systemic pattern library (step 522), the OEM server 108 performs a failure report operation (step 528). The sensor data analyzed for comparison is typically the transformed FFT data. Alternatively, the sensor data is a single sample of raw data (i.e., the sensor signals prior to signal processing) or a time-series set of data. The time-series set of data contains data sets that correspond to a point of time in a time line. When the time-series set of data is used, the last data set (i.e., the last point of data in the time line) is used to select a possible failure pattern as a hypothesis. The hypothesis is compared to the other elements of the time-series set using an appropriate statistical tool to determine if the hypothesis is the likely cause of failure.

[0049] The failure report operation (step 528) includes generating an action alert, generating a report, transmitting the action alert to selected maintenance individuals or to an enterprise asset management, an enterprise resource planning program, or any other maintenance management software operated by the party responsible for maintenance. The report is added to a machine-specific database. The action alert is provided to the party responsible for maintenance of the machine 300 so that appropriate action may be taken. The action alert includes a machine identification, a time stamp, an identification of the component that is likely to fail or that has failed, an estimated time of failure, and a recommended action (i.e., replace, align, check, clean, etc.) The report added to the machine-specific database includes the action alert information and a portion of the sensor data for long term machine monitoring (e.g., historical data to see changes over time).

[0050] If there is no systemic pattern match, the sensor data is compared with known component patterns (step 524). If the sensor data matches a component pattern (step 526), the failure report operation (step 528) is performed. If there is no match, a component ID is assigned and transmitted to the directory server 112 in level 206 (step 530). The component ID is a reference number uniquely describing a machine component, such as a ball bearing, motor or gearbox, etc.. When a match and diagnosis is returned to the OEM server 108, the pattern and diagnosis is added to the component pattern library for use in matching future events.

[0051] Turning now to figure 5c, the directory server 112 searches for OEM servers using the same component with the same component ID sent by the OEM server 108 in level 204 (i.e., the requesting OEM server) (step 540). If a component ID matches (step 542), the directory server 112 sends the server ID of one of the OEM servers with a matching component ID. The requesting OEM server and OEM server with a matching component ID establish a peer-to-peer connection and the data is sent to the OEM server with matching component ID for analysis (step 546). The OEM server with matching component ID compares the sensor data with the system and component pattern libraries (step 548). If there is a match (step 550), the OEM server with matching component ID transmits the diagnosis and component pattern associated with the sensor data to the requesting OEM server 108 in level 204 (step 552). The requesting OEM server 108 receives the information and performs the failure report operation (step 528).

[0052] If there is no match between the sensor data and the OEM server 108 with matching component ID, steps 540 to 550 are repeated with other OEM servers 108 with matching component ID until either a match occurs or no further OEM servers 108 with matching component IDs are found. Alternatively, peer-to-peer connections are established with several OEM servers with matching component IDs so that the OEM servers can perform the sensor data comparison in parallel. If no further OEM servers with matching component IDs are found (i.e., the sensor data does not match any known patterns), the directory server 112 informs the requesting OEM server 108 and establishes a connection with expert network server 114 and transmits the sensor data to the expert network server 114 (step 544).

[0053] Turning now to FIG. 5d, the expert network server 114 receives the sensor data and determines which experts to use. The expert network server 114 identifies a lead expert from a group of experts that will become responsible for solving the problem and establishes a work session with the lead expert (step 560). The group of experts is identified by matching the expertise of the experts with the type of machine 300 that the detector 302 is monitoring. The lead expert is selected based upon a list of criteria. The list of criteria includes availability of the expert, cost, and urgency of the matter. For example, if the

diagnosis must be started immediately, then the group of experts may be narrowed down to those experts that are in an appropriate time zone to start the project (e.g., if the machine problem occurred in the middle of the night in the United States, the lead expert may be chosen from the group of experts residing in that part of the world where the working day is just starting).

[0054] Once the lead expert is identified and agrees to accept the work session, the lead expert analyses the data and identifies specialists to solve the problem (step 564). The specialists work together sharing the same information in a collaborative environment to solve the problem (step 564). The collaborative environment allows the specialists to work together from remote locations. The collaborative environment is a network that provides the specialists and experts with shared access to sensor and machine data, shared access to pattern libraries, document sharing, secure (and non-secure) communications, and the ability to track individual contributions. The communications between the specialists can be voice, video, e-mail, instant messaging, co-browsing, etc. If the specialists chosen are unable to solve the problem (step 566), the lead expert selects other specialists to see if they are able to solve the problem and step 564 is repeated. The lead expert and selected specialists continue to work on the problem until the problem is solved.

[0055] Once the problem is solved, the lead expert validates the solution and determines a failure diagnostic description for placing in the database of the OEM server 108 in level 204 (step 568). The system and component patterns and diagnosis are transmitted to the OEM server 108 in level 204 (step 570). In an alternative embodiment, the system and component patterns are transmitted to all of the OEM servers that have a component ID matching the component ID sent by the requesting OEM server.

[0056] A system and method for a remote multi-level, scalable diagnosis of devices has been described. The foregoing description of various embodiments of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise embodiments disclosed. Numerous modifications or variations are possible in light of the above teachings. The embodiments discussed were chosen and described to provide the best illustration of the principles of the invention and its

practical application to thereby enable one of ordinary skill in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. All such modifications and variations are within the scope of the invention as determined by the appended claims when interpreted in accordance with the breadth to which they are fairly, legally, and equitably entitled.

TO THE OFFICE OF THE SECRETARY OF DEFENSE